REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
28-11-2017	Final Report	1-Dec-2016 - 31-Aug-2017

4. TITLE AND SUBTITLE
Final Report: Identifying Key Actors in Heterogeneous Networks

5a. CONTRACT NUMBER
W911NF-17-1-0009

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Arizona State University
ORSPA
P.O. Box 876011
Tempe, AZ 85287 -6011

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
68299-LS.1

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Elisa Bienenstock
UU	UU	UU	UU		19b. TELEPHONE NUMBER
					602-496-0927

RPPR Final Report
as of 15-Dec-2017

Agency Code:

Proposal Number: 68299LS **Agreement Number: W911NF-17-1-0009**
INVESTIGATOR(S):

Name: Ph.D Elisa Bienenstock
Email: ebienens@asu.edu
Phone Number: 6024960927
Principal: Y

Organization: **Arizona State University**
Address: ORSPA, Tempe, AZ 852876011
Country: USA
DUNS Number: 943360412 EIN: 860196696
Report Date: 30-Nov-2017 Date Received: 28-Nov-2017
Final Report for Period Beginning 01-Dec-2016 and Ending 31-Aug-2017
Title: Identifying Key Actors in Heterogeneous Networks
Begin Performance Period: 01-Dec-2016 **End Performance Period:** 31-Aug-2017
Report Term: 0-Other
Submitted By: Ph.D Elisa Bienenstock Email: ebienens@asu.edu
Phone: (602) 496-0927
Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: **STEM Participants:**

Major Goals: The specific aim of Identifying Key Actors in Heterogeneous Networks (W911NF-17-0009) was to advance and reconcile the formal representations of social relations and coalition formation from social network analysis (SNA) and game theory (GT) to improve accuracy for detecting significant or “powerful” actors within a total actor space when both resource and structural constraints must be taken into account. Our premise was that many critical strategic and tactical situations of interest to the Department of Defense (DOD) present social situations that are outside the scope and violate the assumptions of existing formal social science models. SNA by its very construction focuses on dyadic relations and standard SNA metrics are focused only on structure and do not take into account vertex attributes. And while there has been good work on inferring “roles” from network metrics, these studies do not treat such roles as resources that must coalesce in order to achieve a desired outcome. Cooperative game theory (CGT) models of coalition formation are based on two limiting assumptions: that actors who cooperate do so only to gain the best individual advantage and that all actors can transact with any other. Both of these assumptions are unreasonable in many operational settings. This effort focused on synthesizing SNA and CGT to develop a model that simultaniously considers an actor's intrinsic value to a coalition along with positional factors. For this effort the goal was to develope an approach to address this scenario as a demonstration of a new approach for synthesizing social network analysis and game theory. The ultimate goal of this research agenda is to generalize this approach to support more nuanced analyses of empirical data applicable to a broader range of real-world conditions.

Accomplishments: Identifying Key Actors in Heterogeneous Networks (W911NF-17-0009) produced an entirely unique approach to rank actor influence in a network where structure and role are both factors. The final product includes an easy to use and interpret metric, the beyond dyads (BD) value for ranking vertex “value", which accounts for both the structural constraints modeled by Social Network Analysis (SNA) and the assessment of actors in contributing to collaborative coalitions modeled by cooperative game theory (CGT). The synthetic methodology amalgamated existing constructs from SNA and CG T so that each would complement and overcome the limitations of the other. The new metric presents a major advance in conceptualizing the “actor space” to provide a sizable advantage for identifying and ranking significant actors within the total actor space when both resource and structural constraints must be taken into account, a highly probable situation in many operational contexts. Our initial analysis using this method to evaluate simulated data revealed that the metric behaved as predicted: when structure was absent, the method produced rankings similar to the game-theoretic solutions and when structure was present, the method produced rankings similar to network metrics.

Training Opportunities: Nothing to Report

RPPR Final Report
as of 15-Dec-2017

Results Dissemination: Nothing to Report

Honors and Awards: Nothing to Report

Protocol Activity Status:

Technology Transfer: Nothing to Report

PARTICIPANTS:

Participant Type: Consultant
Participant: Michael Salwen
Person Months Worked: 6.00
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Funding Support:

FINAL REPORT

Identifying Key Actors in Heterogeneous Networks

(W911NF-17-0009)

Organization
ARIZONA STATE UNIVERSITY
Tempe, 852876011, USA

Principal Investigator:
Dr. Bienenstock, Elisa

Forward

The specific aim of *Identifying Key Actors in Heterogeneous Networks* (W911NF-17-0009) was to advance and reconcile the formal representations of social relations and coalition formation from social network analysis (SNA) and game theory (GT) to improve accuracy for detecting significant or "powerful" actors within a total actor space when both resource and structural constraints must be taken into account. This project produced a new analytical capability forged by integrating the structural constraints that can be modeled by SNA with the assessment of actors contributing to collaborative coalitions modeled by cooperative game theory (CGT). Our premise was that many critical strategic and tactical situations of interest to the Department of Defense (DOD) present social situations that are outside the scope and violate the assumptions of these existing social science models. SNA by its very construction focuses on dyadic relations and standard SNA metrics are focused only on structure and do not take into account vertex attributes. And while there has been good work on inferring "roles" from network metrics, these studies do not treat such roles as resources that must coalesce in order to achieve a desired outcome. CGT models of coalition formation are based on two limiting assumptions: that actors who cooperate do so only to gain the best individual advantage and that all actors can transact with any other. Both of these assumptions are unreasonable in many operational settings. The situations we focused on for this effort entailed a network of actors with different skills seeking to form an alliance with complementary, though not necessarily directly connected, alters to achieve an advantageous outcome at the expense of others. The approach developed to address this scenario is intended as a demonstration of a new approach for synthesizing social network analysis and game theory. The ultimate goal of this research agenda is to generalize this approach to support more nuanced analyses of empirical data applicable to a broader range of real-world conditions.

The project just completed produced a new metric that we refer to as the Beyond Dyads (BD) value, which accounts for both the structural constraints modeled by SNA and the assessment of actors in contributing to collaborative coalitions modeled by cooperative game theory (Shubik and Shapley 1972). The synthetic methodology amalgamated existing constructs from SNA and GT so that each would complement and overcome the limitations of the other. The new metric presents a major advance in conceptualizing the "actor space" to provide a sizable advantage for identifying and ranking significant actors within the total actor space when both resource and structural constraints must be taken into account, a highly probable situation in many operational contexts. Although the motivation for this project was to develop a better approach for identifying key actors in opposition to US interests, the models and methods developed will apply to a broad range of scenarios regarding the formation or dissolution of coalitions that impact US interests and policy objectives.

The primary objective of this report is to describe the BD algorithm. In the next section we provide a theoretically motivated example that conveys the essence of our approach in a manner

that allows the reader to understand the problem we are addressing and the means we have conceived to solve it without requiring deep study of the mathematical model. We use this example as a reference as it provides an intuitive illustration of the limitations of GT and SNA taken singly, and the potential benefit of an integrated approach. This introduction is followed by a detailed description of how that BD metric is calculated. We then return to the illustrative example to describe how we implemented and evaluated this new metric.

Table of Contents

Statement of the Problem Studied

Theoretically Motivated Example

To illustrate our approach consider the following *very simple* example. In the network diagram given in Figure 1, edges denote structural connections among actors and shapes and colors denote certain roles or resources required to perform a collective goal. We imagine that actors are aware of each other and can act collectively only if they are directly connected to one another or if they act in concert with other actors that provide the necessary linkages. That is, we suppose

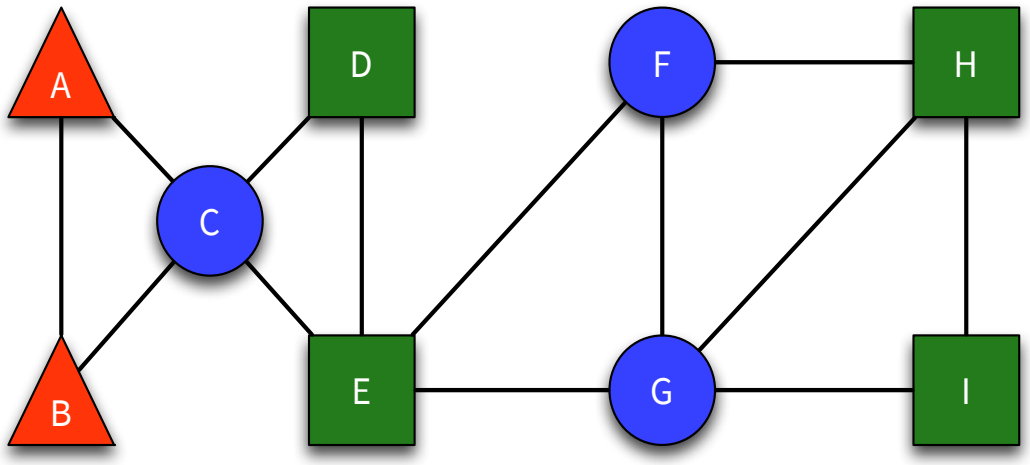

Figure 1. Actor space for the example, including structural ties and attributes

that no coalition can form among actors unless a path exists within the coalition linking each member to every other. Additionally, we require that for a coalition to accomplish the goal, it must contain at least one actor possessing each of the three required resources (triangle, circle, and square). An *intuitive* inspection of Figure 1, leads most people to identify vertex *C* as the preeminently significant actor, but SNA and GT solutions move the focus to others, because structural and role constraints taken together require rethinking how power or significance is evaluated within the space of actors. Were we to focus only on the structural constraints of the network and use the most frequently used standard network metrics, significance appears to be broadly distributed throughout the network (see Table 1 below). Using a game theory solution, the Shapley value, to gauge the power of actors based only on attributes and without accounting for structural constraints, we identify the two triangular vertices as most significant because of their relative scarcity. (Note that for this example, we have used a characteristic function for each coalition equal to the number of disjoint triangle-circle-square (T-C-S) triples contained in the coalition. Since there are at least two of each shape, the grand coalition has value 2.)

Our approach aims to identify key actors in networks of this type that are too large and complex for humans to develop an intuition without analytical tools. In actuality, the methods we will

produce will operate on complex datasets that do not reveal an intuitively obvious solution, and will allow for nuanced analyses to distinguish among “imperfect” coalitions based upon their degree of deficiency with respect to roles or resources and structural proximity. The example provided above was intended merely to illustrate the limitations of using either GT or SNA methods in isolation. In the following section we describe our approach to creating a new metric that considers both structure and role.

Summary of the Most Important Results

Methodology for computing Beyond Dyads values

To emphasize that our context is an actor space with a distribution of attributes or roles over it where interaction among actors is constrained by a network topology, we will denote this actor network as G to emphasize that it is a graph as well as a set of actors. We are borrowing ideas from SNA and cooperative game theory CGT, though our assumptions are distinct from those made within either of those traditions. In particular, we assume that actors will cooperate toward a common goal irrespective of personal gain. This stands in stark contrast to the assumption in CGT that actors cooperate for maximal personal gain.

A key influence from CGT on our approach is how actor valuations derive from coalition valuations. In CGT this is often accomplished through the computation of the Shapley value on the basis of some *characteristic* function, which is a real-valued set function defined on subsets, often called coalitions, of the actor space. In this formulation, an actor's value is the accumulation of the value contributed to each coalition, which in turn is measured by assessing the difference in the characteristic function value of the coalition with and without the actor as a member.

Our approach will be to build a set function that can be used similarly. This metric should incorporate both the number or, more generally, weight of attributes possessed by a subset of the actor space and a measure of its connectivity.

Before delving into the construction of the set function, we must consider a central barrier to this type of approach. Functions such as the Shapley value are usually computed over the power set of the actor space. As the number of subsets of a set of n elements equals 2^n, this number explodes quickly. Most methods for managing this exponential explosion reduce this space of subsets by replacing the entire power set with some intelligently or randomly selected number of subsets from it. We will adopt the former strategy, and use our knowledge of the attributes distributed over the actor space to selectively produce a set of subsets (more accurately, subgraphs) that will be the basis for our set valuations, and from those our vertex valuations.

Problem context

The problem domain we are studying is one in which we suppose that there is an ensemble $\{a_1, a_2, \ldots, a_N\}$ of N attributes distributed across the actor space and that one actor possessing each attribute is required to form a coalition capable of accomplishing some goal. We also assume that these actors are vertices in a social network and that there may be actors in the space which possess none of these attributes, but which may provide key connections within the network.

For now, we assume that each actor may possess at most one of the special attributes. In the future, we may wish to modify this constraint to allow actors to possess several of the traits necessary to accomplishing the collective goal. We may also wish to generalize to the case where we require n_i actors possessing attribute $a_i, i = 1,2,\ldots,N$ to achieve the goal, and, within this latter context, to the case where N may equal 1, that is, where there is no distinction among the attributes, but merely a constraint that a certain number of actors possessing *the* key attribute must cooperate to achieve the goal.

Subgraph selection

The first technical requirement of any method approaching the problem as we have is to cut down the sample space of actor subsets. At the outset of this project, we assumed that for this STIR phase, we would do no better than to sample subsets of the actor space randomly, and to base our metric upon such a random sample. However, upon a bit more reflection, it became clear that the requirement that coalitions of members possessing certain key attributes be formed to accomplish the desired could be leveraged to intelligently rather than randomly sample the subset space. The method we arrived at is detailed immediately below.

We start the process of selecting subgraphs by forming tuples of actors representing distinct attributes. For this phase, we required tuples to be *complete*, that is, to contain one actor representing each of the N attributes. We can consider relaxing this requirement in the future, thus expanding the set of sampled subsets to those containing some minimum number (or, more generally, weight) of attributes, and our algorithms are coded to accommodate such a change. We denote by $\{A_0, A_1,\ldots,A_N\}$ the partition of the elements of G by attribute, where elements of A_i possess attribute a_i and the subset A_0 contains elements of G possessing none of the required attributes. With this notation in place, we can formally define a tuple T as an element of the Cartesian product set $A_1 \times A_2 \times \ldots \times A_N$. Note that A_0 is not part of this product set. Since we are concerned with connectivity as well as attributes, *we create tuples only within connected components of* G, not across them.

Once the set of tuples is in hand, we create subgraphs in the following way. For each dyad of vertices in a tuple T, we calculate all minimal distance—by which we mean minimal length or, more generally, minimal edge weight—paths joining the elements of the dyad. Having obtained the set of paths for each dyad (note, this set may contain a unique minimal path or several such paths), we then take the Cartesian product of these path sets. We obtain a subgraph from each element of this product set of paths by taking the subgraph of G induced by the set of vertices lying on the paths that comprise this element. Note that because minimal paths are not necessarily unique, a tuple can give rise to more than one subgraph. It is equally true that the same subgraph can be induced by different tuples. In the end, we include one copy of each

distinct subgraph created by this process in the collection of subgraphs that will form the basis for vertex valuations.

We can illustrate the process with the following very simple example. Suppose we have a graph (or at least a piece of graph that contains all the relevant information we need for a given tuple of vertices) that looks like this

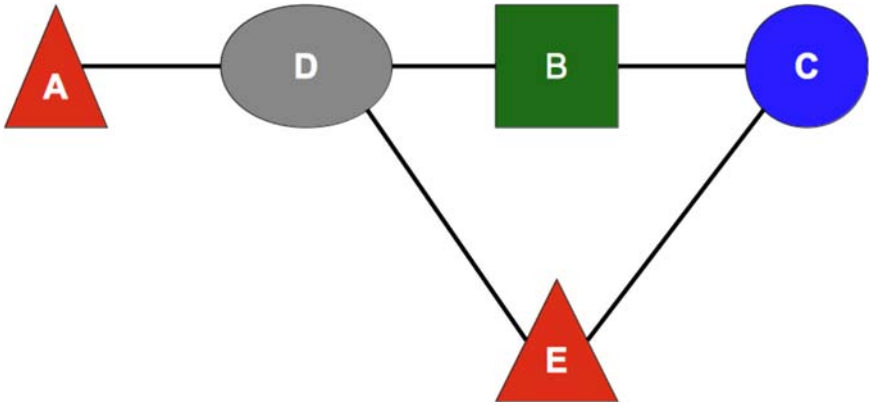

Figure 2

and that we wish to derive the subgraphs from the tuple (*A*, *B*, *C*). There are unique minimal paths between *A* and *B* and between *B* and *C*; however, between *A* and *C* there are two minimal paths: *A*—*D*—*B*—*C* and *A*—*D*—*E*—*C*. This gives rise to two distinct subgraphs: $S_1 = \{A, B, C, D\}$ and $S_2 = \{A, B, C, D, E\}$. Note that all the edges that already exist between vertices are included in the generated subgraphs. In this example, the edge between *B* and *D* is preserved in the second subgraph, even though it does not figure in any of the minimal paths used to generate the subgraph.

This little example highlights some key points and raises a few questions. First, the second path from *A* to *C* engenders a second subgraph only because this path is also minimal. If it were not because, for example, there was an additional vertex between *E* and *C* or because the weights of the edges on this path were greater than those on the path through *B*, then there would be no S_2, even if there was some second path, perhaps not minimal, but also perhaps not too long or weighty either. We may wish to broaden our sample of subgraphs by including paths between dyads taken from the tuple that are either (a) minimal or (b) not minimal, but below some distance (length or weight) threshold, and which have intervening vertices that are *disjoint* from the tuple.

How we actually value subgraphs and, therefore, vertices, is covered in the next section, but a couple of high-level points illustrated by our simple example can be made now. Note first that

the existence of the second path linking *A* to *C* means that some residual value remains in this dyad even after *B* is removed from subgraph S_2. Without this second path there would be no S_2, and *A* and *C* would never have the opportunity to realize this value. It should be noted that *B* also gets extra value from the existence of this second path between *A* and *C* as it receives much, but not all, of the value of S_2, which is in addition to the value it receives from S_1.

Next, let's focus on S_1. When we delete *B* from this subgraph, the resulting isolated key vertices yield no value *because they lack connection*. The value thus gained by *B* equals the total value of S_1. But, when we delete *A, C,* or *D* from S_1, the value of the remaining dyad containing *B*, while not zero, is diminished. Therefore, some fraction of S_1's value is granted to *A, C* and *D*. It should be clear that the sum of these four vertex values deriving from subgraph S_1 exceeds the value of S_1. This is not, in itself, problematic for our purposes of determining relative valuations among vertices, but it is in contrast to the behavior of valuations like the Shapley value.

Subgraph and vertex valuation

The top-level idea here is that we will use tuples of vertices possessing distinct key attributes to value the graphs within which they lie. For example, if red, green and blue are the three required attributes, for a given graph we form all (red, green, blue) triples (where, as before, tuples are created within not across connected components), assign a value to each triple as a function of attribute weight and minimal path distances between dyads, and take the maximum of these tuple values as the value for the graph itself.

The details of how these values are computed follows, but before moving to that let's take note of a detail glossed over in the brief top-level description. A choice of what constitutes a valuable tuple of vertices must be made. The implication made above is that we are only interested in complete tuples, that is, tuples containing a representative of each attribute class. At the moment, our default method is to select subgraphs for the sample that are generated only by complete tuples. We may wish to modify this to include subgraphs generated by incomplete tuples of attribute representatives if they reach some attribute threshold. This threshold can be expressed in terms of the number of attributes or the fraction of total attribute weight represented or both. Our algorithms are written to accept either or both of these types of thresholds.

We require a set of weights $\{w_1, w_2, \ldots, w_N\}$ on our set of N attributes $\{a_1, a_2, \ldots, a_N\}$ such that $w_i \in (0,1]$ for all i and $\sum_i w_i = 1$. Within these immediate constraints, our method is indifferent to how these weights are found. For the present, we will assign weights based upon the frequency of attributes in the population of actors.

Suppose that there exist n_i members of the actor space possessing attribute a_i for $i = 1,2,\ldots,N$. Since scarcity should imply value, we assign weights to each attribute as

$$w_i = \frac{\frac{1}{n_i}}{\sum_{j=1}^{n} \frac{1}{n_j}}, i = 1,2,\ldots,N.$$

In the following, it will be necessary to associate these weights to vertices possessing one of the attributes. We denote by $w_a(x)$ the value of the attribute possessed by actor x; that is, $w_a(x) = w_i$ if x possesses attribute a_i. If x possesses none of the key attributes, we set $w_a(x) = 0$. Since we will be interested in connections between actors possessing the important attributes, this zero value does not factor into our calculations, but is included to define a function on the complete actor space. This attribute weight and the path distance between dyads will be the basis of our tuple valuations.

To value a graph we collect all tuples and value each in the following way. For each dyad (x,y) in the tuple, we denote by $d(x,y)$ the minimal path distance between x and y, where distance is the minimal path length or, more generally, the minimal sum of path edge weights within the subgraph. We then compute a dyad value $v(x,y)$ as

$$v(x,y) = \frac{w_a(x) + w_a(y)}{d(x,y)^{\alpha}},$$

where α is a tunable parameter in the $(0,1]$ interval. Our starting value for α is $\frac{1}{3}$ (cube root) and we expect that values should lie in the $\left[\frac{1}{4},\frac{1}{2}\right]$ interval (4th root to square root). We can consider α a path-distance penalty parameter, with larger values exacting a larger penalty on the dyad value for the minimal path distance between the vertices.

We will aggregate these dyad values over the tuple T by summing, normalizing and scaling. Summing is rather straightforward. To normalize, note that any dyad value achieves its maximum when there is a distance 1 path between the vertices (this is clearly independent of the choice of α since 1^{α} for any α). In this case, $v(x,y) = w_a(x) + w_a(y)$. If this maximal situation exists for all dyads of the tuple, then the sum of all dyad values is

$$\sum_{(x,y)} v(x,y) = \sum_{(x,y)} \Big(w_a(x) + w_a(y) \Big),$$

where the sum is taken over all dyads of T. This last sum is equal to $\big(N-1\big)\sum_{x \in T} w_a(x)$, that is, the sum of the attribute values over the elements of T multiplied by the length of T minus 1. To see this, simply note that each $x \in T$ is paired once with every other, so that it is in exactly $N-1$ dyads. Thus $w_a(x)$ appears in the original sum exactly $N-1$ times. Putting this all together, we can normalize the sum of the dyad values for a tuple T by dividing by this maximum possible value, that is,

$$\frac{\sum_{(x,y)} v(x,y)}{\big(N-1\big)\sum_{x \in T} w_a(x)}.$$

Note that this value lies in the $(0,1]$ interval; however, it achieves the maximum value of 1 whenever each dyad in T is connected by a path with distance 1. But this is true whether T contains a representative of every attribute class or just some of them. For this reason, our last step in valuing T is to scale the term above by the total attribute weight represented by T. If T contains a representative of every attribute class, this value will be 1. Otherwise, this value will lie in $(0,1)$. With a slight abuse of notation we will denote this final value for T as $v(T)$ and we get a nice simple expression:

$$\frac{\sum_{(x,y)} v(x,y)}{\big(N-1\big)\sum_{x \in T} w_a(x)} \cdot \sum_{x \in T} w_a(x) = \frac{\sum_{(x,y)} v(x,y)}{\big(N-1\big)}.$$

So, in words, our tuple of length N has a value equal to the sum of all dyad values divided by $N-1$. The dyad values are the sum of the attribute values represented by the vertices of the dyad, diminished by a divisor that is a function of the graph distance between the vertices.

Note that $v(T)$ equals $\frac{\overline{v(x,y)}}{2}$, where we are denoting by $\overline{v(x,y)}$ the average dyad value. This observation allows us to think of $v(T)$ as the sum we would obtain over all elements of T were we to distribute to each of these vertices one-half of the average dyad value. We take the maximum value of $v(T)$ over all tuples T formed from the graph and assign that value to the graph. Vertices are valued by removing them from each of the sample subgraphs in which they lie and accumulating the deltas of the set values that result.

As part of this effort we also investigated avenues for adjusting these values to give a scoring advantage to vertices that reduce a high-value set to a moderate-value set over those that would yield a similar difference but between a moderate-value set and a low-value set. For example, in terms of non-transformed values, a vertex that reduces the value of a high-value set from 0.8 to 0.6 gains the same yield (0.2) that a vertex that reduces the value of a moderate-value set from 0.5 to 0.3. The effect of distorting the final value nonlinearly can be seen in the Figure 3, where the nonlinear function is given by $f(x) = \frac{1+e^{\gamma+\beta}}{1+e^{\gamma+\beta x}}$. In this plot, $\beta = 5$ and $\gamma = 3.5$. If we transform these values by the function shown above, the first vertex would gain $f(0.8) - f(0.6) = 0.28$, and the second would gain $f(0.5) - f(0.3) = 0.20$.

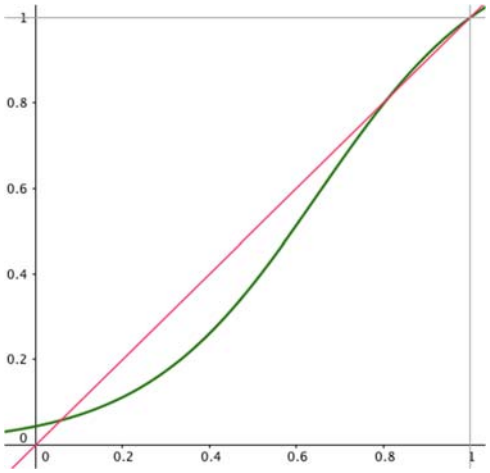

Figure 3

Generation of BD Value

The BD algorithm was programmed in Python to accept input as raw comma-delimited data or in the graphML format. Data represents an edge list and associated vertex attribute file. For any graph alternative, "colorings" can be saved in a graphML file as attributes. Gray is the default color we have used to represent vertices with no intrinsic value to the coalition.

Instructions for computing BD values are detailed in the Jupyter Notebook which will be included as part of the software package we are preparing. The algorithm accommodates the simultaneous processing of multiple (Freeman 1978; Bonacich, 1987) on the same graph. The output currently includes centrality metrics and BD scores for all specified colorings.

Table 1 is a summary of the BD values for each vertex in the network represented in Figure 1. For this illustration we also included values for the GT Shapley Value and Core (Shapley & Shubik 1971), and the SNA Key Player Solution (Borgatti 2006). Note that GT solutions identify vertices A and B as the most advantaged, while the game theory solutions identify vertex E as the

most powerful. The BD solution successfully identified vertex C as the most powerful or advantaged with regard to access to resources given network structures.

Table 1

		Game Theory Solutions		Network Solutions					
Vertex Label	Original Coloring	Shapley Value	Core	Betweeness	Closeness	Degree	Eigenvector	Key Player	BD Values
A	red	0.56	1	0	0.806	0.692	0.465	0	1.751
B	red	0.56	1	0	0.806	0.692	0.465	0	1.751
C	blue	0.18	0	2.769	1.152	1.385	1.003	0	4.307
D	green	0.08	0	0	1.008	0.692	0.779	0	0.119
E	green	0.08	0	3.692	1.344	1.385	1.458	1	1.073
F	blue	0.18	0	0.577	1.075	1.038	1.309	0	0
G	blue	0.18	0	1.846	1.152	1.385	1.512	0	0
H	green	0.08	0	0.115	0.849	1.038	1.162	0	0
I	green	0.08	0	0	0.806	0.692	0.847	0	0

Metrics Evaluation

To evaluate the metric required that we compare it to existing approaches under a range of conditions. The first step was to create an assortment of colored graphs and access how well the BD algorithm did at identifying advantaged positions compared to the conventional metrics. The inherent challenge to evaluating the value and accuracy of this new metric is that, with the exception of very simple examples like the network in Figure 1, the “correct” answer is unknowable. If we could ascertain the answer in any other way, this algorithm would not be needed.

We met this challenge by constructing families of colored graphs with an array of structural properties that should theoretically generate rankings approximating either known network metric values or game theory values. We then assessed whether the BD metric results seemed reasonable given these general expectations. For this project, our objective was to confirm our intuition that the metric was performing in a manner that weighted both structure and role. This was neither comprehensive nor conclusive. A true test the validity and usefulness of the algorithm will require applying the logic presented here to a broader range of simulated networks, and complementing that with applying the algorithm to analyze empirical datasets. At this stage in the development of the algorithm, our objective was to apply the metric to a theoretically informative selection of example networks to deepen our understanding of how the algorithm was performing. We anticipate that the logic of this approach will ultimately be the

basis of an approach to measure metric validity. Currently, the standard we hoped to meet was to demonstrate that the metric output was reasonable given the structural properties of the network.

We considered three of types of 21-vertex graphs: random (Erdos & Renyi 1960), preferential attachment (Barabasi and Albert 1999) and small world (Watts & Strogatz 1998) each of varying attribute densities ranging from equal distribution to highly skewed distribution of the three colors needed to complete a coalition (red, blue and green) and a fourth color (gray) that indicated vertices not possessing any key attribute. Game theoretical solutions produce identical rankings for each coloring scheme regardless of network structure. The network solutions each provide identical rankings for each network regardless of colorings. Therefore, in the absence of structure, the BD value should match game theoretic predictions, and as structure becomes more salient, the BD value should diverge and become more similar to social network metric rankings. For instance, a completely connected network, or, to a lesser extent, a random network, should approximate the values generated by the Shapley value. In contrast, the vertex rankings of a highly structured graph, such as a small world network, with a random distribution of coloring, should yield rankings more similar to centrality scores.

For this project we investigated the network in Figure 1 and one 21-vertex network of each type with varying coloring schemes. We also ran the algorithm for networks of 100 vertices, primarily to verify that the algorithm could support networks of that size. Once initial values for the baseline network were calculated, we investigated the impact on BD scores of swapping the colorings of specific pairs of vertices in the same structure to investigate how that impacted the rankings (Figure 4). Similarly, we experimented with removing or adding edges to investigate how small structural changes impact rankings. The BD metric, as expected produced rankings similar to those of the Shapley Value when networks were either very dense or random. For highly structured small world and scale free networks, the BD metric produced rankings similar to centrality, especially when colors were randomly and evenly distributed.

Table 2 includes the BD values for the original illustrative network for three alternative colorings, demonstrating how the distribution of colors within the structure alter the power hierarchy.

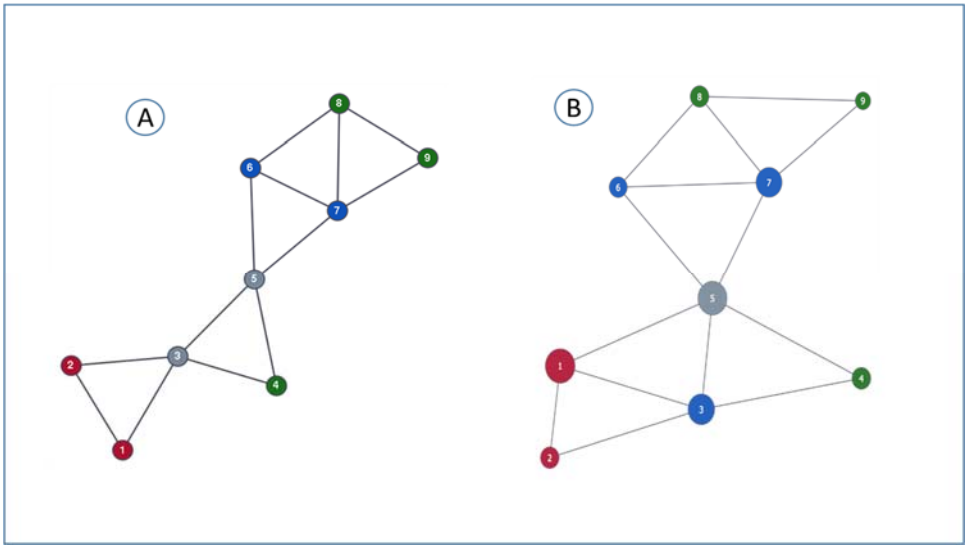

Figure 4. Alternative colorings for identical structures reveal the impact of minor modifications in structure to the BD metric. The addition of vertices or edges reveal how small structural perturbations impact BD rankings (Graph B).

Table 2

		Alternative Colorings			BD Values			
Vertex Label	Original Coloring	G-3	G-5	G 3&5	Original	G-3	G-5	G-3&5
A	red	red	red	red	1.751	0.943	1.475	0.697
B	red	red	red	red	1.751	0.943	1.475	0.697
C	blue	gray	blue	gray	4.307	1.886	3.978	1.394
D	green	green	green	green	0.119	0	0.81	0.474
E	green	green	gray	gray	1.073	3.152	0.505	1.394
F	blue	blue	blue	blue	0	0.89	0.084	1.034
G	blue	blue	blue	blue	0	1.187	0.252	2.058
H	green	green	green	green	0	0	0.252	0.751
I	green	green	green	green	0	0	0.168	0.501

Conclusion

Identifying Key Actors in Heterogeneous Networks (W911NF-17-0009), a six month Short-Term Innovative Research project concluding on August 31, 2017, produced an entirely unique approach to rank actor influence in a network where structure and role are both factors. The final product includes an easy to use and interpret algorithm for ranking vertex "value." Our initial analysis using this method to evaluate simulated data revealed that the metric behaved as

predicted: when structure was absent, the method produced rankings similar to the game-theoretic solutions and when structure was present, the method produced rankings similar to network metrics.

Bibliography

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *science*, *286*(5439), 509-512.

Bonacich, P. (1987). Power and centrality: A family of measures. *American journal of sociology*, *92*(5), 1170-1182.

Borgatti, S. P. (2006). Identifying sets of key players in a social network. *Computational & Mathematical Organization Theory*, *12*(1), 21-34.

Erdos, P., & Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, *5*(1), 17-60.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social networks*, *1*(3), 215-239.

Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games*, *2*(28), 307-317.

Shapley, L. S., & Shubik, M. (1971). The assignment game I: The core. *International Journal of game theory*, *1*(1), 111-130.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *nature*, *393*(6684), 440-442.